Numerical integration methods for large-scale biophysical simulations

Edoardo Milotti [a,*] Alessio Del Fabbro [a] Roberto Chignola [b]

[a] *Dipartimento di Fisica dell'Università di Trieste, and I.N.F.N. – Sezione di Trieste, Via Valerio 2 – I-34127 Trieste, Italy*

[b] *Dipartimento di Biotecnologie dell'Università di Verona, and I.N.F.N. – Sezione di Trieste, Strada Le Grazie 15 - CV1, I-37134 Verona, Italy*

Abstract

Simulations of biophysical systems inevitably include steps that correspond to time integrations of ordinary differential equations. These equations are often related to enzyme action in the synthesis and destruction of molecular species, and in the regulation of transport of molecules into and out of the cell or cellular compartments. Enzyme action is almost invariably modeled with the quasi-steady-state Michaelis-Menten formula or its close relative, the Hill formula: this description leads to systems of equations that may be stiff and hard to integrate, and poses unusual computational challenges in simulations where a smooth evolution is interrupted by the discrete events that mark the cells' lives. This is the case of a numerical model (Virtual Biophysics Lab - VBL) that we are developing to simulate the growth of three-dimensional tumor cell aggregates (spheroids). The program must be robust and stable, and must be able to accept frequent changes in the underlying theoretical model: here we study the applicability of known integration methods to this unusual context and we describe the results of numerical tests in situations similar to those found in actual simulations.

Key words: biomolecular simulations, stiff systems in biophysics, timescales in biophysics
PACS: 87.10.Ed, 87.17.Aa, 87.18.Vf, 87.18.Nq

* corresponding author: Edoardo Milotti
  Email addresses: `milotti@ts.infn.it` (Edoardo Milotti),
`Alessio.Delfabbro@ts.infn.it` (Alessio Del Fabbro), `chignola@ts.infn.it`
(Roberto Chignola).

1 Introduction

Most mathematical models of biochemical and biophysical processes in cells are described by nonlinear differential systems that cannot be handled with analytical methods and require numerical integrations [1,2]. However the high speed of computers and the sophisticated computational methods that are available today are powerful tools that allow the numerical exploration of these exceedingly interesting dynamical systems. This also suggests that eventually the biophysical models will no longer be analytic, but mostly computational; and indeed we are now developing a program that simulates tumor spheroids (VBL, Virtual Biophysics Lab) [3,4,5,6], which includes a reduced but still quite complex description of the biochemistry of individual cells, plus many diffusion processes that bring oxygen and nutrients into cells and metabolites into the environment.

In a numerical model like VBL, each cell is described by a reduced metabolic network, and by other mechanisms that include both discrete deterministic and stochastic events. The description is thus mixed, with smooth evolutions interspersed with discrete steps. The exchange of molecules with the surrounding environment means that transport into and out of cells is closely linked with diffusion processes that involve the whole cluster of cells, and finally lead to a very large set of (time) differential equations. Simulation steps between discrete events require the integration of nonlinear differential equations that describe the individual cell's clockwork, and the integration of the diffusion equations. These integrations are carried out under widely different conditions, in a changing environment, and for this reason they need algorithms that are both unconditionally stable and free from unwanted artifacts. These conditions are not always fulfilled in the existing literature, and we feel that a detailed understanding of the underlying mathematical and computational bases may be important not just for us, but for other workers in the field of systems biology as well.

Eventually the whole simulation program must run smoothly, and it must be free of stability problems. Thus it is very important to ensure that step by step integrations are stable and do not bring the biochemical variables into unphysical regions (e.g., no concentration must ever become negative).

2 Biophysical models

Biophysical models are not arbitrary dynamical systems: indeed the phenomena of life are characterized by remarkable stability properties, as most of them display either homeostasis or (nearly) stable limit cycles. At the basic

reaction level, many processes obey a straightforward enzyme kinetics, regulated by the well-known Michaelis-Menten equation [1], or by a variant that applies to cooperative processes, the Hill equation [7].

Here we start with a special case that displays generic features shared by many other processes, namely the transport of glucose into and out of cells and its conversion into glucose-6-phosphate (G6P; this is part of the reduced metabolic model incorporated in our simulator of cell metabolism, growth and proliferation [4,5]). We begin with the equations for a single cell in a stable environment:

$$
V_{in}\frac{d\,[G_{in}]}{dt} = \frac{v_{\mathrm{max},1}\,[G_{extra}]}{K_1 + [G_{extra}]} - \frac{v_{\mathrm{max},1}\,[G_{in}]}{K_1 + [G_{in}]}
$$
$$
- \frac{v_{\mathrm{max},2}\,[G_{in}]^2}{(K_2 + [G_{in}])\,(K_a + [G_{in}])}
$$
$$
- \frac{v_{\mathrm{max},22}\,[G_{in}]^2}{(K_{22} + [G_{in}])\,(K_a + [G_{in}])} \tag{1}
$$
$$
V_{extra}\frac{d\,[G_{extra}]}{dt} = -\frac{v_{\mathrm{max},1}\,[G_{extra}]}{K_1 + [G_{extra}]} + \frac{v_{\mathrm{max},1}\,[G_{in}]}{K_1 + [G_{in}]}
$$
$$
+ D\frac{[G_{env}] - [G_{extra}]}{\Delta}S \tag{2}
$$

Here the dynamical variables are the glucose concentration inside the cell $[G_{in}]$, and the concentration in the extracellular space $[G_{extra}]$; the environmental glucose concentration $[G_{env}]$ is fixed and is one of the model parameters. The introduction of the extracellular space may look like a useless nuisance in this case, but it is justified on two different, and both important, grounds. The extracellular space is absolutely needed in multicellular systems, because in that case it becomes the scaffolding that allows the diffusion of many substances in closely bound cell clusters (and corresponds to an actual biological entity, the free space available in the extracellular matrix [8,9,10]), and a consistent treatment requires its extension also towards the environment; secondly, the extracellular space, even in this simple case, acts as the cell-environment interface, and corresponds to the buffer region actually observed in diffusion measurements in tumor spheroids [11].

We assume a spherical cell of radius $r \approx 5\mu$m, so that the surface area is $S = 4\pi r^2 \approx 3 \cdot 10^{-10}$m^2, and the volume is $V_{in} = (4/3)\pi r^3 \approx 5 \cdot 10^{-16}$m^3. The first two terms on the rhs of equation (1) correspond to the transport of glucose from the extracellular space into the cell and the reverse process of transport from the inside of the cell back to the extracellular space (enacted by the reversible GLUT transporters, see [12,13] and references therein); since this transport process is facilitated [12,13] we describe it with two Michaelis-Menten terms, where the maximum transport speed $v_{\mathrm{max},1}$ is proportional

to the cell's surface [4,12,13], and K_1 is an experimental parameter [4].The other two terms correspond to a cascade of two Michaelis-Menten processes that involve the enzymes glucokinase and hexokinase, and is regulated by parameters that do not depend on cell size [14].

The second equation (eq. (2)), in addition to the GLUT-mediated transport terms, includes a standard diffusion term that corresponds to the diffusion of glucose from the surrounding environment into the extracellular space. This term is a discretization of glucose flux across the surface S (which is approximately the area of the cell-extracellular space interface); the distance used to compute the concentration gradient is the thickness of the extracellular layer, $\Delta \approx 0.2\mu$m.

Although equations like (1) and (2) are usually cast in terms of concentrations, we prefer to use masses, as this makes total mass conservation stand out clearly:

$$
\begin{aligned}
\frac{dm_{in}}{dt} =& \frac{v_{\max,1}m_{extra}}{V_{extra}K_1 + m_{extra}} - \frac{v_{\max,1}m_{in}}{V_{in}K_1 + m_{in}} \\
& - \frac{v_{\max,2}m_{in}^2}{(V_{in}K_2 + m_{in})(V_{in}K_a + m_{in})} \\
& - \frac{v_{\max,22}m_{in}^2}{(V_{in}K_{22} + m_{in})(V_{in}K_a + m_{in})} \\
\frac{dm_{extra}}{dt} =& - \frac{v_{\max,1}m_{extra}}{V_{extra}K_1 + m_{extra}} + \frac{v_{\max,1}m_{in}}{V_{in}K_1 + m_{in}} \\
& + \frac{DS}{\Delta}\left([G_{env}] - \frac{m_{extra}}{V_{extra}}\right)
\end{aligned}
$$

(3)

(4)

We postpone the listing of model parameters to section 4, where we summarize the results of numerical tests.

3 Integration methods

There is a broad and comprehensive literature on integration methods, and at first it may seem that in large-scale biophysical simulations the choice is only limited by the required accuracy and by the available computing power, however it is not so. In a simulation like VBL [4,5] there are many cells and the number of equations is quite large: our aim is to simulate at least one million cells, which corresponds to a spherical cell cluster with a diameter of one millimeter. In VBL we make several simplifying assumptions and one of them is that each cell has its own extracellular space: this means that in such a system – for each molecular species which diffuses in the cell cluster – there

are about two million local concentration variables (for each cell there is an intracellular and an extracellular concentration) and a corresponding number of equations.

The simulation includes random events as well (such as the partitioning of organelles at the time of cell division), so that numerical inaccuracies are absorbed in this much larger randomness and turn out to be much less important than they are in strictly deterministic systems.

Moreover, the life of cells is marked by discrete events: this discreteness produces endless transient responses in the deterministic part of the simulation algorithm, and the most evident discrete process – cell division – also changes the number of variables and equations.

Last but not least, there is a wide range of volumes (cells and extracellular spaces), and this leads to very stiff systems of equations: in fact while the metabolic processes can be quite slow (with characteristic times of the order of thousands of seconds), a single extracellular space has a very small volume and the time it takes to fill or empty this volume is very short (of the order of 10–100 μs). Thus the characteristic times span eight orders of magnitude, and the system is quite stiff. This means that any adopted integration method must be very stable, and that it should be able to overlook very short transients that do not mean much from a biophysical point of view, and it should also accept long time steps without crashing or producing unphysical results.

All this proves quite challenging, as it poses conflicting requests on the integration algorithms. The usual considerations on stability [15,16,17] suggest that implicit methods should perform better than explicit methods, and this also means that a robust algorithm must sacrifice speed, as implicit methods require several function evaluations.

Multistep methods [16,17,18] may appear to be a smart choice, because they can incorporate information from previous simulation steps and thus reduce the number of required function evaluations: however numerical tests performed in anticipation of the present work have shown that explicit Adams-Bashforth multistep methods are highly unstable in this context, and that the fourth-order Adams-Moulton implicit multistep method does not fare much better as it seems to be very sensitive to the quality of the algorithm initialization steps (that must necessarily be performed with another integration algorithm), and unfortunately the discrete events lead to a constant flow of equation initialization processes.

Finally we have decided to include the following methods in the tests summarized in the next section:

- the default integrator of *Mathematica* [19], as a benchmark algorithm;

- the implicit Euler method;
- the implicit trapezoidal method (which corresponds to the second-order Adams-Moulton algorithm).

Although it cannot be included as it is in a simulation program, we have decided to choose the default setting of the *Mathematica* instruction `NDSolve` as a benchmark since it incorporates a well-tuned mixture of standard algorithms not included in our very short list [19], and thus provides a qualitative guide to possible better choices. Here simplicity is an important bonus, since the integration algorithm must be incorporated in a larger structure and the biophysical model may change frequently, as new biochemical paths are incorporated: however we have not included the simple explicit Euler method in the list, because it is well-known to fail badly in the case of stiff equations, unless the step size is exceedingly small (and yet this method is widely used in many similar contexts, see, e.g, [20]). The many important Runge-Kutta (RK) methods are also missing, because the stable implicit RK methods are not easily adopted for inclusion in a simulation program like VBL (or other similar numerical stepping schemes), and also because the sophisticated integrator of *Mathematica* [19] actually includes a choice of RK methods – both explicit and implicit – and its performance may suggest corrections to our selection of algorithms.

Since stability – and not precision – is the main feature that an integrator must have in this context, it would be important to set the general problem in a form such that the standard theorems on stability can be applied [16]. Unfortunately it is impossible to specify exactly the operating conditions of the simulator, especially because the number of equations is variable – as cells proliferate – and because the theoretical model must be open to changes as our understanding of the biophysical and biochemical processes evolves. However there are a couple of general considerations that can help: the first is that the systems of equations must be autonomous, time cannot appear explicitly. The second is that the equations describe the evolution of masses and concentrations, i.e. quantities that must be non-negative, and this means that the generic structure of the equations for non-negative quantities x_k must be

$$\frac{dx_k}{dt} = A(x_1, \ldots, x_N; t) - B(x_1, \ldots, x_N; t)x_k \tag{5}$$

where A and B are non-negative functions that express respectively production and consumption/destruction of the quantity x_k. The consumption/destruction term must be proportional to x_k to ensure the continued non-negativity of the solution (if there were not such a proportionality, then the derivative could be nonvanishing and negative for values of x_k close to zero, and thus lead to negative x_k). Indeed the proportionality is often related to Michaelis-Menten

6

terms with a generic structure $x_k/(K_m + x_k)$. A differential system such as (5) can be locally approximated by the simpler equations

$$\frac{dx_k}{dt} = A_0 - B_0 x_k \qquad (6)$$

which shows that A-stable integration methods should normally be sufficient in this complex context [16].

4 Numerical results on the test model of section 2

We have chosen parameter values derived from experiment or from fits of experimental data and they are listed in table 1; all parameters are further explained and referenced in [4,5], except D [21] and Δ. The approximate value of Δ is derived from the estimate that the extracellular matrix amounts to about 20% of the total volume of human tissue [10] and assuming a free space fraction of about 50% [8] (we note that this may actually be an overestimate): then we obtain $\Delta \approx 0.5 \, (0.25 V_{in}) \, /S \approx 0.2 \mu$m.

Using the values in the table we can find the equilibrium values that correspond to $dm_{in}/dt = 0$ and $dm_{extra}/dt = 0$, i.e., $[G_{in}] \approx 0.043$ kg m^{-3} and $[G_{extra}] \approx 0.90$ kg m^{-3}: this means that because of conversion into G6P, the glucose level in this isolated cell is considerably lower than in the surrounding environment.

As explained above, the system of equations (3) and (4) is very stiff, and this can be easily visualized setting an off-equilibrium initial glucose concentration inside the cell. Some of the glucose is quickly converted into G6P, but part of it either seeps out or enters the cell, and changes the concentration in the extracellular space. Since the extracellular volume is very small it reacts very quickly to this sudden change, and exchanges glucose with the environment, exhibiting a sharp transient. The approximate duration of this transient can be estimated expanding equation (4) for very low values of the masses: since $V_{in} \gg V_{extra}$,

$$\frac{dm_{extra}}{dt} \simeq - \left(\frac{v_{\max,1}}{V_{extra} K_1} + \frac{DS}{V_{extra} \Delta} \right) m_{extra} + \frac{DS}{\Delta} [G_{env}], \qquad (7)$$

with the parameters of table 1, this gives $m_{extra}(t) \sim A + B \exp(-t/\tau)$, with $\tau \approx 60 \mu$s. It is also important to note that this very short time is entirely due to the diffusion term in equation (7): without the diffusion term the characteristic time τ would grow to more than 27 s.

parameter	value	
r	$5 \cdot 10^{-6}$ m	cell radius
$S = 4\pi r^2$	$3 \cdot 10^{-10}$ m	surface area
$V_{in} = \frac{4\pi}{3} r^3$	$5 \cdot 10^{-16}$ m^2	cell volume
Δ	$0.2 \cdot 10^{-6}$ m	thickness of extracellular space around cell
$V_{extra} = S\Delta$	$6 \cdot 10^{-17}$ m^3	extracellular volume
$v_{max,1} \approx 2 \cdot 10^{-9} S$	$6 \cdot 10^{-19}$ kg s^{-1}	maximum speed of GLUT-mediated glucose transport
$v_{max,2}$	$1.2 \cdot 10^{-19}$ kg s^{-1}	maximum rate of glucokinase activity
$v_{max,22}$	$1.2 \cdot 10^{-18}$ kg s^{-1}	maximum rate of hexokinase activity
K_1	0.27024 kg m^{-3}	Michaelis-Menten constant of glucose transport
K_2	1.8 kg m^{-3}	Michaelis-Menten constant for glucokinase
K_{22}	$1.8 \cdot 10^{-2}$ kg m^{-3}	Michaelis-Menten constant for hexokinase
K_a	$5.4 \cdot 10^{-2}$ kg m^{-3}	constant for the homeostatic loop controlling glucokinase and hexokinase activity
D	$7 \cdot 10^{-10}$ m^2 s^{-1}	diffusion constant of glucose in water
$[G_{env}]$	0.9 kg m^{-3}	glucose concentration in standard nourishing solutions

Table 1
Parameters used in the numerical tests with one isolated cell. Parameters are explained and referenced in [4,5], except D [21] and Δ (see discussion in the main text).

The nearly exponential transient is clearly visible in a plot of the mass flow into the extracellular space

$$\left. \frac{dm_{extra}}{dt} \right|_{inflow} = \frac{DS}{\Delta} \left([G_{env}] - \frac{m_{extra}}{V_{extra}} \right) \tag{8}$$

which is shown in figure 1 for the initial conditions $m_{in} = 0.1 \, [G_{env}] \, V_{in}$ and $m_{extra} = 0.1 \, [G_{env}] \, V_{extra}$. Because of this fast transient, many standard algorithms fail when integration is carried out on long time intervals with comparatively long time steps: rather surprisingly, *Mathematica* itself fails badly (at least with the standard settings for NDSolve) even though it includes a stiffness detection procedure [19]. This failure becomes slowly apparent as the integration is carried out on longer and longer intervals; figure 2) shows the concentration $[G_{in}]$ vs. time for a 1000 seconds interval and initial condi-
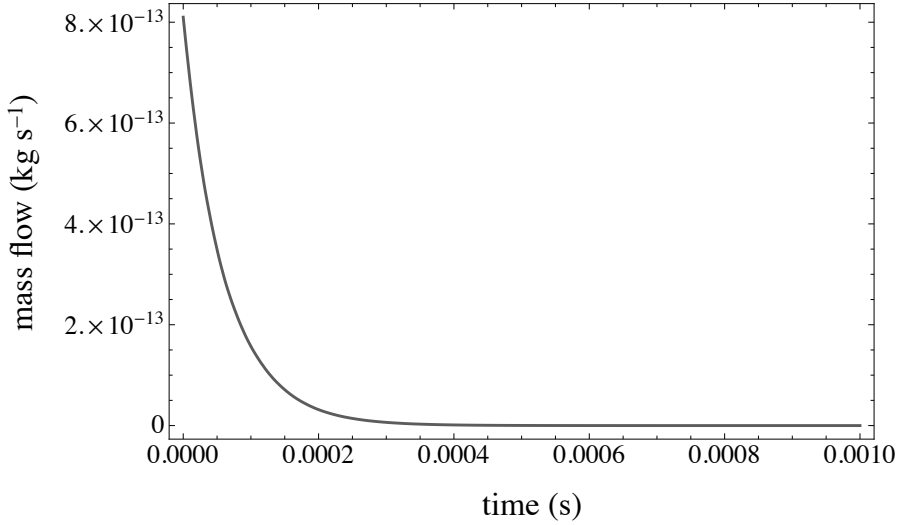
Fig. 1. Mass flow into the extracellular space (mass flow (kg s^{-1}) vs. time (s)) obtained from the numerical integration of equations (3) and (4) with *Mathematica*. A direct fit of this nearly exponential curve yields a decay time $\tau \approx 62\mu$s, very close to the estimate in the main text.

tions $m_{in} = 0.9\,[G_{env}]\,V_{in}$ and $m_{extra} = 0.1\,[G_{env}]\,V_{extra}$: instead of a smooth decay towards the equilibrium value the *Mathematica* solution shows some unexpected undulations. The companion plot for the concentration in the extracellular space $[G_{extra}]$ (not shown) displays a large, unphysical peak close to the origin (this peak is about 100 times larger than the environmental concentration). A similar integration performed on a longer interval (10000 s) returns an even worse result, as it produces negative and unphysical values for the concentration $[G_{in}]$ (see figure 3).

A careful analysis of the curve in figure 1 reveals that the fast transient is very close to an exponential, and therefore we expect the standard results on A-stability of integration methods [16] to be applicable to this and to similar nonlinear differential systems. In particular, we know that both the implicit Euler and the implicit trapezoidal method are unconditionally A-stable. Indeed the tests performed show that both methods converge even when the internal machinery of *Mathematica* – which uses a full array of sophisticated methods – is foiled by this stiff system. And although it is known that the accuracy of the implicit trapezoidal method (IT) is higher than that of the implicit Euler algorithm (IE) [16], in these tests IE fares better, since the IT method shoots below the equilibrium value of $[G_{in}]$ before swinging – very slowly back (see figure 4), and displays unwanted oscillations of $[G_{extra}]$ that very gradually die out. The IT method also alternates steps above and below the equilibrium value of $[G_{extra}]$, and these oscillations depend on the step size and resemble those observed in a closely related method, the Crank-Nicolson
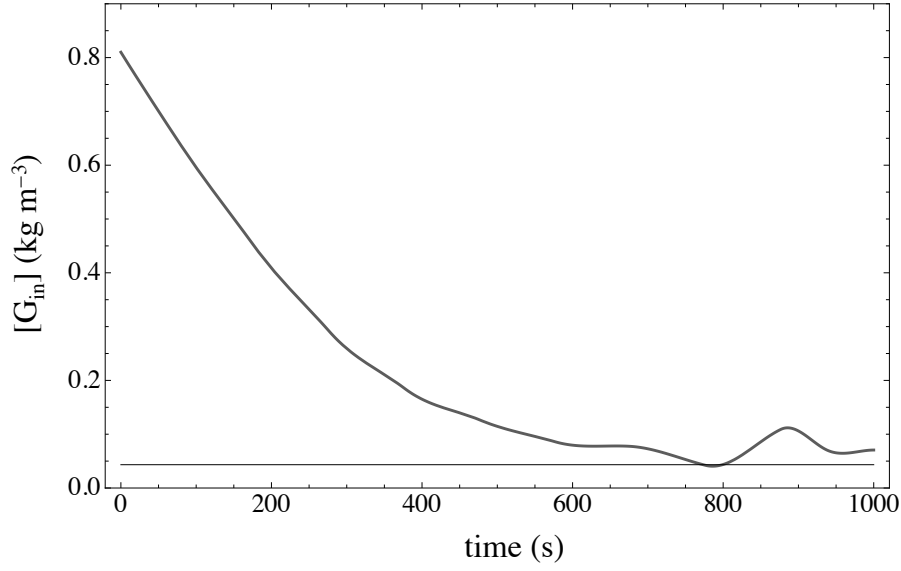
9

Fig. 2. Plot of the concentration $[G_{in}]$ (kg m^{-3}) vs. time (s) obtained from the numerical integration of equations (3) and (4) with *Mathematica* for a 1000 seconds long time interval. The gray horizontal line shows the equilibrium level. The internal stepping procedure (with the default values) fails and produces unexpected undulations.
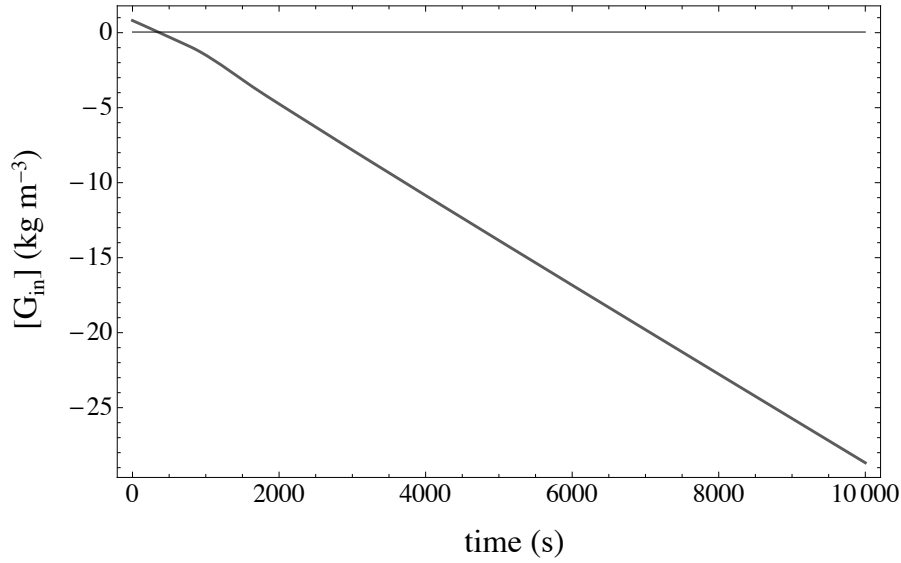

Fig. 3. Plot of the concentration $[G_{in}]$ (kg m^{-3}) vs. time (s) obtained from the numerical integration of equations (3) and (4) with *Mathematica* for a 10000 seconds long time interval. The gray horizontal line shows the equilibrium level. The internal stepping procedure (with the default values) fails badly and leads to large, negative (unphysical) values of the concentration $[G_{in}]$.
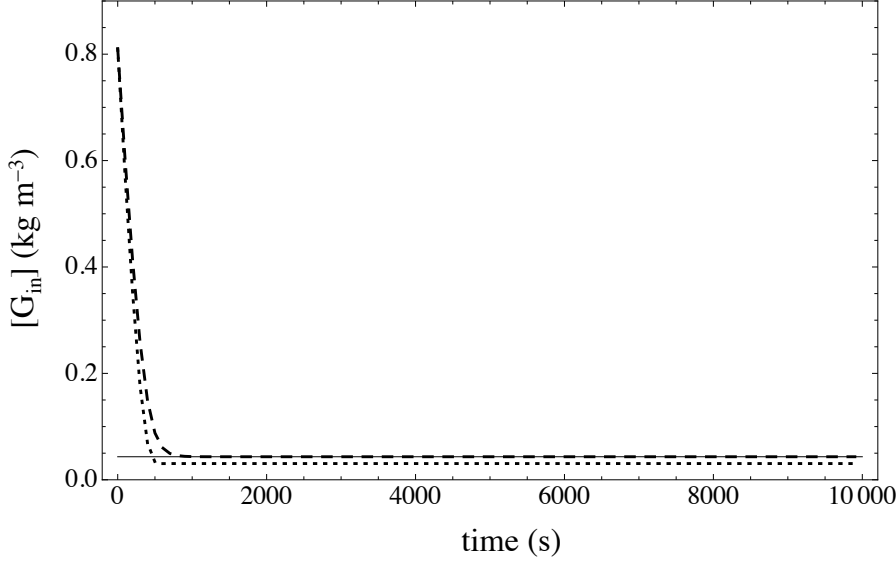
Fig. 4. Plot of the concentration $[G_{in}]$ (kg m^{-3}) vs. time (s) obtained from the numerical integration of equations (3) and (4) with the implicit Euler method (dashed curve) and with the implicit trapezoidal method (dotted curve) for a 10000 seconds long time interval; in each case the step size is 100 s. The gray horizontal line shows the equilibrium level: the implicit Euler method reaches this equilibrium value without unwanted oscillations, while the implicit trapezoidal method shoots below the line and afterwards climbs very slowly back to the equilibrium value. We find that this climb depends on the actual stepsize, and it is faster for shorter time steps. Neither algorithm produces unphysical values (negative concentrations).

method for partial differential equations [22,23].

5 Diffusion in cell clusters

In the previous section we have considered just one cell, but as explained above, we wish to simulate large clusters of cells, and this means that we have to deal with large systems of coupled nonlinear differential equations. We begin with a very small cluster of just two cells, where glucose dynamics is described by the following equations for the masses

$$
\frac{dm_{in}^{(A)}}{dt} = \frac{v_{\max,1}^{(A)} m_{extra}^{(A)}}{V_{extra}^{(A)} K_1 + m_{extra}^{(A)}} - \frac{v_{\max,1}^{(A)} m_{in}^{(A)}}{V_{in}^{(A)} K_1 + m_{in}^{(A)}} +
$$
$$
\frac{v_{\max,1}^{(AB)} m_{extra}^{(AB)}}{V_{extra}^{(AB)} K_1 + m_{extra}^{(AB)}} - \frac{v_{\max,1}^{(AB)} m_{in}^{(A)}}{V_{in}^{(A)} K_1 + m_{in}^{(A)}}
$$

11

$$-\frac{v_{\max,2}\left(m_{in}^{(A)}\right)^2}{\left(V_{in}^{(A)}K_2 + m_{in}^{(A)}\right)\left(V_{in}^{(A)}K_a + m_{in}^{(A)}\right)}$$

$$-\frac{v_{\max,22}\left(m_{in}^{(A)}\right)^2}{\left(V_{in}^{(A)}K_{22} + m_{in}^{(A)}\right)\left(V_{in}^{(A)}K_a + m_{in}^{(A)}\right)} \tag{9}$$

$$\frac{dm_{in}^{(B)}}{dt} = \frac{v_{\max,1}^{(B)}m_{extra}^{(B)}}{V_{extra}^{(B)}K_1 + m_{extra}^{(B)}} - \frac{v_{\max,1}^{(B)}m_{in}^{(B)}}{V_{in}^{(B)}K_1 + m_{in}^{(B)}}$$

$$+\frac{v_{\max,1}^{(AB)}m_{extra}^{(AB)}}{V_{in}^{(AB)}K_1 + m_{extra}^{(AB)}} - \frac{v_{\max,1}^{(AB)}m_{in}^{(B)}}{V_{in}^{(B)}K_1 + m_{in}^{(B)}}$$

$$-\frac{v_{\max,2}\left(m_{in}^{(B)}\right)^2}{\left(V_{in}^{(B)}K_2 + m_{in}^{(B)}\right)\left(V_{in}^{(B)}K_a + m_{in}^{(B)}\right)}$$

$$-\frac{v_{\max,22}\left(m_{in}^{(B)}\right)^2}{\left(V_{in}^{(B)}K_{22} + m_{in}^{(B)}\right)\left(V_{in}^{(B)}K_a + m_{in}^{(B)}\right)} \tag{10}$$

$$\frac{dm_{extra}^{(A)}}{dt} = -\frac{v_{\max,1}^{(A)}m_{extra}^{(A)}}{V_{extra}^{(A)}K_1 + m_{extra}^{(A)}} + \frac{v_{\max,1}^{(A)}m_{in}^{(A)}}{V_{in}^{(A)}K_1 + m_{in}^{(A)}}$$

$$+\frac{DS^{(A)}}{\Delta}\left([G_{env}] - \frac{m_{extra}^{(A)}}{V_{extra}^{(A)}}\right) \tag{11}$$

$$\frac{dm_{extra}^{(B)}}{dt} = -\frac{v_{\max,1}^{(B)}m_{extra}^{(B)}}{V_{extra}^{(B)}K_1 + m_{extra}^{(B)}} + \frac{v_{\max,1}^{(B)}m_{in}^{(B)}}{V_{in}^{(B)}K_1 + m_{in}^{(B)}}$$

$$+\frac{DS^{(B)}}{\Delta}\left([G_{env}] - \frac{m_{extra}^{(B)}}{V_{extra}^{(B)}}\right) \tag{12}$$

$$\frac{dm_{extra}^{(AB)}}{dt} = -2\frac{v_{\max,1}^{(AB)}m_{extra}^{(AB)}}{V_{extra}^{(AB)}K_1 + m_{extra}^{(AB)}} + \frac{v_{\max,1}^{(AB)}m_{in}^{(A)}}{V_{in}^{(A)}K_1 + m_{in}^{(A)}} + \frac{v_{\max,1}^{(AB)}m_{in}^{(B)}}{V_{in}^{(B)}K_1 + m_{in}^{(B)}}$$

$$+\frac{DS^{(L)}}{\Delta_L}\left([G_{env}] - \frac{m_{extra}^{(AB)}}{V_{extra}^{(AB)}}\right) \tag{13}$$

where the superscripts (A) and (B) denote the two cells, and the parameters in the numerical tests are a straightforward extension of those used for a single cell and are listed in table 2.

The geometric structure of the two cells is depicted in figure 5: this structure approximates the shape of two cells just after mitosis, and we distinguish six different regions

- the two hemispherical cells;
- two hemispherical extracellular spaces that are the interface between the cells and the environment;

parameter	value	
$r^{(A)} = r^{(B)} = r$	$5 \cdot 10^{-6}$ m	cell radius
Δ	$0.2 \cdot 10^{-6}$ m	thickness of extracellular space around cell
$\Delta_L = r$	$5 \cdot 10^{-6}$ m	effective distance for calculation of flow into and out of the extracellular space between cells
$S = 4\pi r^2$	$3 \cdot 10^{-10}$ m	total surface area
$S^{(A)} = S^{(B)} = S/2$	$1.5 \cdot 10^{-10}$ m	surface area of each hemisphere
$S^{(AB)} = \pi r^2$	$7.9 \cdot 10^{-11}$ m	contact area between cells
$S^{(L)}$	$1.5 \cdot 10^{-10}$ m	surface area of each hemisphere
$V_{in}^{(A)} = V_{in}^{(B)} = \frac{2\pi}{3} r^3$	$2.5 \cdot 10^{-16}$ m^2	cell volume
$V_{extra}^{(A)} = V_{extra}^{(B)} = 2\pi r^2 \Delta$	$3 \cdot 10^{-17}$ m^2	volume of each hemispherical extracellular space
$V_{extra}^{(AB)} = \pi r^2 \Delta$	$1.6 \cdot 10^{-17}$ m^2	volume of extracellular space between cells
$v_{max,1}^{(A)} = v_{max,1}^{(B)}$	$3 \cdot 10^{-19}$ kg s^{-1}	maximum speed of GLUT-mediated glucose transport between each cell and surrounding hemispherical extracellular space
$v_{max,1}^{(AB)}$	$1.6 \cdot 10^{-19}$ kg s^{-1}	maximum speed of GLUT-mediated glucose transport between each cell and the extracellular space between cells
$v_{max,2}$	$1.2 \cdot 10^{-19}$ kg s^{-1}	maximum rate of glucokinase activity
$v_{max,22}$	$1.2 \cdot 10^{-18}$ kg s^{-1}	maximum rate of hexokinase activity
K_1	0.27024 kg m^{-3}	Michaelis-Menten constant of glucose transport
K_2	1.8 kg m^{-3}	Michaelis-Menten constant for glucokinase
K_{22}	$1.8 \cdot 10^{-2}$ kg m^{-3}	Michaelis-Menten constant for hexokinase
K_a	$5.4 \cdot 10^{-2}$ kg m^{-3}	constant for the homeostatic loop controlling glucokinase and hexokinase activity
D	$7 \cdot 10^{-10}$ m^2 s^{-1}	diffusion constant of glucose in water
$[G_{env}]$	0.9 kg m^{-3}	glucose concentration in standard nourishing solutions

Table 2

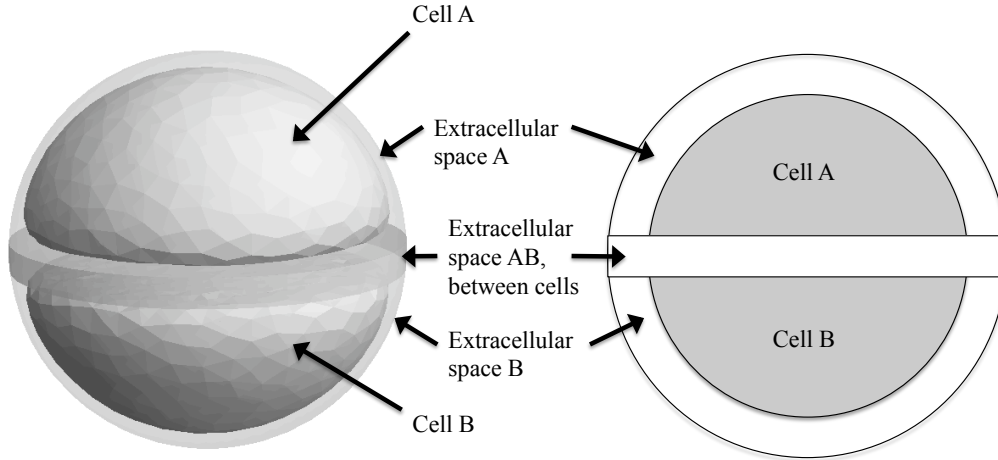Parameters used in the numerical tests with two cells.

Fig. 5. Schematic view of the geometric structure of the pair of daughter cells described in the main text, where it is assumed that just after mitosis the cells are (nearly) hemispherical and are surrounded by two hemispherical extracellular spaces – the interface with the environment – and are separated by one cylindrical extracellular space. Left panel: perspective view; right panel: cross section. For clarity the thickness of the extracellular spaces is exaggerated; the interface between extracellular spaces is actually negligible and is not counted in the equations (9) to (13).

- the cylindrical extracellular space between the cells; this cylindrical space is in contact with the environment only through a narrow belt;
- the (fixed) environment.

This rough division of space contains the main elements of simulations where diffusion plays an important role, and later in this section we shall see how to expand it.

We have carried out extensive numerical tests similar to those in the previous section, and essentially they confirm what we found with a single cell:

- the *Mathematica* solution obtained with the default settings fails for long integration intervals;
- the implicit Euler method and the implicit trapezoidal method both converge to the equilibrium value, but the implicit trapezoidal method displays strong and very undesirable oscillations.

This small system with just two cells can be extended further, with the addition of more cells and their extracellular spaces: this is equivalent to the discrete volume version of a diffusion problem (see, e.g., [24] for a review of volume discretization for the solution of diffusion problems), i.e., we must deal with a large system of equations (here we use concentrations ρ rather than

the masses because the ensuing analysis is slightly easier to follow):

$$V_a \frac{d\rho_a}{dt} = F_a\left(\rho_a; \rho_a\right) + D\sum_{\langle b \rangle}\left(\rho_a - \rho_b\right) g_{ab} \qquad (14)$$

where indexes denote both cells and extracellular spaces, g_{ab} is term related to geometry (it is equivalent to the S/Δ ratio that we met previously), the F_a's describe facilitated transport (for both cells and extracellular spaces) and metabolic processes (cells only), and the diffusion term is actually included only in the case of extracellular spaces. The subscript $\langle b \rangle$ indicates that the sum is performed only on the set of cells b that are adjacent to cell a: since in a simulator like VBL the cells are in arbitrary positions in space [3], and not on the usual cubic lattice, the number of neighbors is random (in a 3D configuration it fluctuates about an average of 12).

We have already remarked in the previous section that the F_a's are slowly varying functions and that the stiffness of the differential system and the algorithmic stability problems are almost entirely due to the diffusion terms: for this reason we concentrate on a reduced version of the differential system (14), i.e.,

$$\frac{d\rho_a}{dt} \approx \frac{D}{V_a}\sum_{\langle b \rangle}\left(\rho_a - \rho_b\right) g_{ab} \qquad (15)$$

It is well-known that differential systems like (15), obtained from the discretization of diffusion problems on regular lattices, can be integrated by implicit methods like the Backward Differentiation Formulas (BDF) or the Crank-Nicolson algorithm (CN) [17,22], and that in those cases both BDF and CN are unconditionally stable [22]. However in simulation programs like VBL, the equations are not discretized on a lattice, and may have a variable number of diffusion terms, as the sum runs over a random number of neighbors: are these implicit integration methods still stable in this more general setting? Fortunately the answer is yes, because it is possible to develop a variation of the standard Von Neumann stability analysis. The argument for BDF goes as follows: we start from the discrete-time version of (15) that corresponds to the BDF iterations

$$\rho_a^{n+1} = \rho_a^n + \Delta t \frac{D}{V_a}\sum_{\langle b \rangle}\left(\rho_b^{n+1} - \rho_a^{n+1}\right) g_{ab} \qquad (16)$$

where Δt is the time step, $\rho_a^n = \rho_a(n\Delta t)$, and we take the test solution $\rho_a^n \simeq A^n \exp\left(i\mathbf{k} \cdot \mathbf{r}_a\right)$ that corresponds to the test solution used in the standard Von Neumann stability analysis, but without the assumption of a regular spatial

lattice. Substituting the test solution in the iteration formulas (16) we find

$$A^{-1} = 1 + \frac{D\Delta t}{V_a} \sum_{\langle b \rangle} g_{ab} \left(1 - \cos \phi_{ab}\right) - i \frac{D\Delta t}{V_a} \sum_{\langle b \rangle} g_{ab} \sin \phi_{ab} \tag{17}$$

where $\phi_{ab} = \mathbf{k} \cdot (\mathbf{r_b} - \mathbf{r_a})$, and since $\sum_{\langle b \rangle} g_{ab} \left(1 - \cos \phi_{ab}\right) \geq 0$ we see that $|A| \leq 1$ and thus the BDF algorithm is unconditionally stable.

The CN iteration formula is

$$\rho_a^{n+1} = \rho_a^n + \Delta t \frac{D}{2V_a} \left[\sum_{\langle b \rangle} \left(\rho_b^{n+1} - \rho_a^{n+1} \right) g_{ab} + \sum_{\langle b \rangle} \left(\rho_b^n - \rho_a^n \right) g_{ab} \right] \tag{18}$$

and we can repeat the steps followed for the BDF algorithm and find

$$A = \frac{1 - \frac{D\Delta t}{V_a} \sum_{\langle b \rangle} g_{ab} \left(1 - \cos \phi_{ab}\right) + i \frac{D\Delta t}{V_a} \sum_{\langle b \rangle} g_{ab} \sin \phi_{ab}}{1 + \frac{D\Delta t}{V_a} \sum_{\langle b \rangle} g_{ab} \left(1 - \cos \phi_{ab}\right) - i \frac{D\Delta t}{V_a} \sum_{\langle b \rangle} g_{ab} \sin \phi_{ab}} \tag{19}$$

and just as before we see that $|A| \leq 1$ and that the CN algorithm is unconditionally stable.

To test the BDF and the CN algorithm in a realistic biophysical setting, we have extended the two-cell model to a 1D string of cells, where we assume that each cell is surrounded by its own extracellular space, and that adjacent extracellular spaces can exchange material by diffusion. This means that the diffusion problem becomes slightly more complex (although the previous stability analysis still holds) and the equations for each cell/extracellular space are

$$V_C \frac{d\rho_C}{dt} = M\left(\rho_C\right) + T\left(\rho_c, \rho_C\right) \tag{20}$$

$$V_c \frac{d\rho_c}{dt} = D \sum_{\langle b \rangle} \left(\rho_b - \rho_c\right) g_{bc} - T\left(\rho_c, \rho_C\right) \tag{21}$$

where

$$M_G\left(\rho_{G,C}\right) = -\frac{v_{\max,2}\rho_{G,C}^2}{\left(K_2 + \rho_{G,C}\right)\left(K_a + \rho_{G,C}\right)}$$
$$-\frac{v_{\max,22}\rho_{G,C}^2}{\left(K_{22} + \rho_{G,C}\right)\left(K_a + \rho_{G,C}\right)} \tag{22}$$

$$T_G\left(\rho_{G,C}, \rho_{G,c}\right) = \frac{v_{\max,1}\rho_{G,c}}{K_1 + \rho_{G,c}} - \frac{v_{\max,1}\rho_{G,C}}{K_1 + \rho_{G,C}} \tag{23}$$
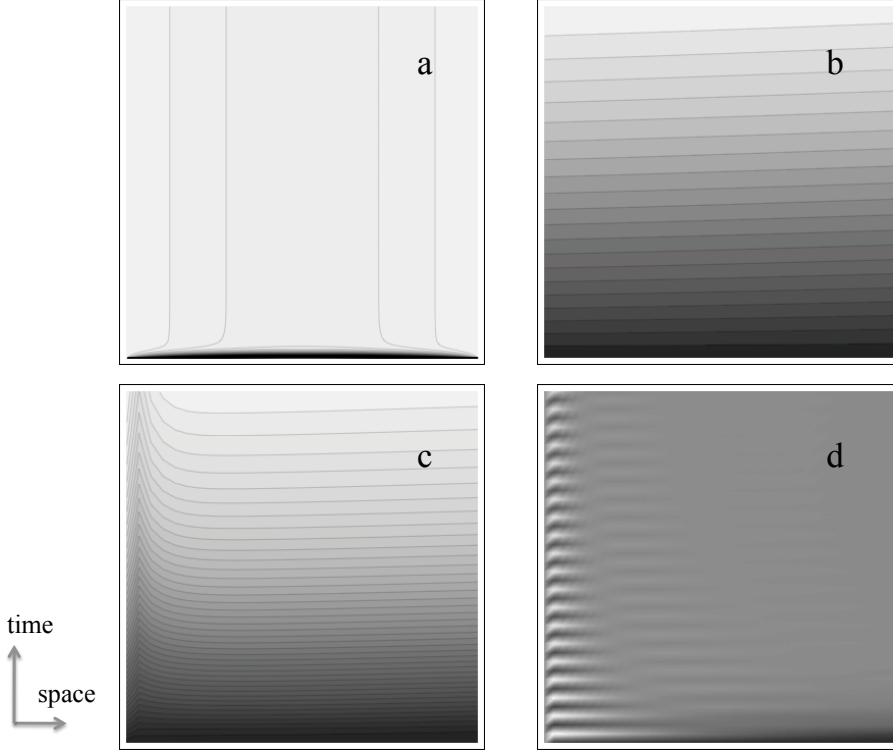
Fig. 6. Contour/density plots that show different aspects of the solutions of equations (20-21) for a string of 100 cells, that spans a spatial range $(-500\mu m, 500\mu m)$. Space runs along the x-axis and time runs upwards along the y-axis with timesteps $\Delta t = 2$ s, and concentration values are also mapped on a gray scale (black correspond to the minimum value, white to maximum). The panels show: **a)** the contour plot of the glucose concentration in the extracellular spaces obtained with the BDF method; **b)** contour plot of the glucose concentrations inside cells, for cells in the reduced spatial range $(-500\mu m, -200\mu m)$, for $t = 0$ s until $t = 60$ s, obtained with the BDF method; **c)** same as **b**, but in this case the numerical solution has been calculated with the CN method: notice that the contour lines are appreciably distorted close to the boundary; **d)** contour plot of the glucose concentrations in the extracellular spaces in the reduced spatial range $(-500\mu m, -200\mu m)$, for $t = 0$ s until $t = 60$ s, obtained with the CN method: notice the banded structure that is due to the oscillatory instability of the CN method.

are the functions that describe the internal metabolic activity (conversion of glucose to G6P) and the transport processes into and out of cells and that we met earlier both in the one-cell and in the two-cell models. The uppercase subscripts denote cells and the lowercase subscripts denote extracellular spaces, and all the model parameters have already been listed in table 2. The space coordinate (cell positions) span the range (x_{min}, x_{max}), with the boundary conditions $u(x_{min}, t) = u(x_{max}, t) = $ constant: thus this is a model with both diffusion of glucose and local absorption and conversion into G6P.

The BDF and the CN algorithm are equivalent to the implicit Euler and to the
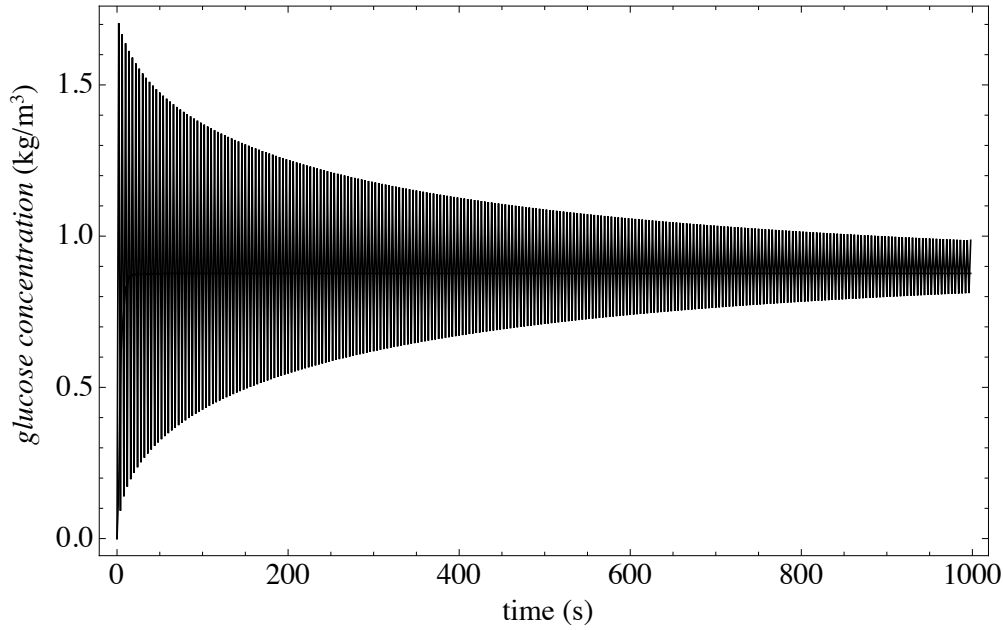
Fig. 7. Plot of the glucose concentration in the leftmost extracellular space in the string of cells described in the text, obtained numerically with the CN method for the same problem of figure 6. In this solution there is a marked high frequency spurious oscillation that depends on the time step and that very slowly fades away.
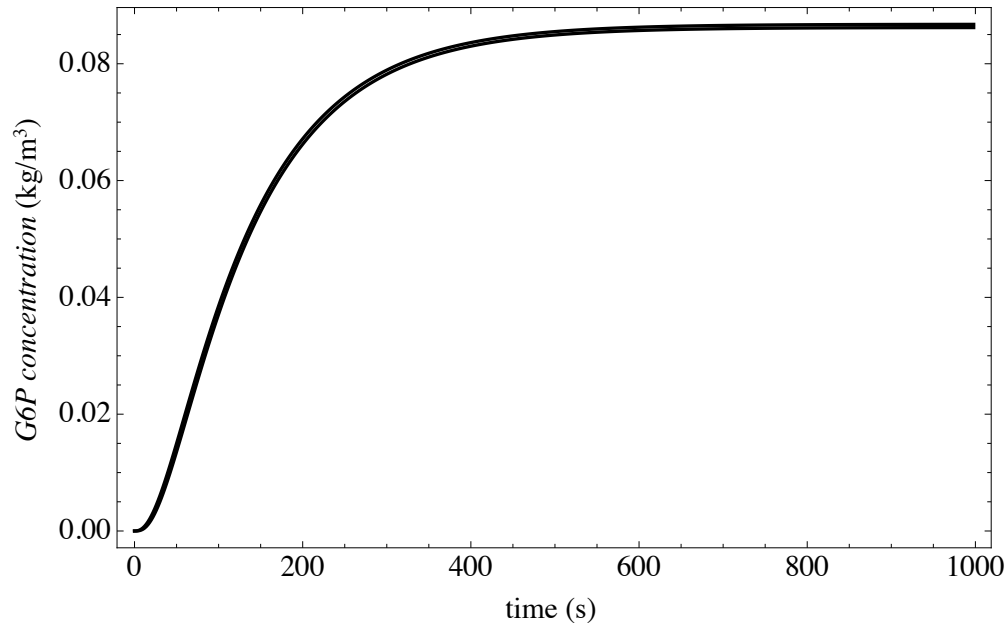

Fig. 8. Plot of G6P concentration at two positions (close to the boundary and in the middle of the string of cells), obtained numerically with the BDF method for the same problem of figure 6, using the differential equation for G6P in [4], solved with the implicit Euler method; the concentration of G6P is almost independent of position along the string and the two curves are indistinguishable on this scale.
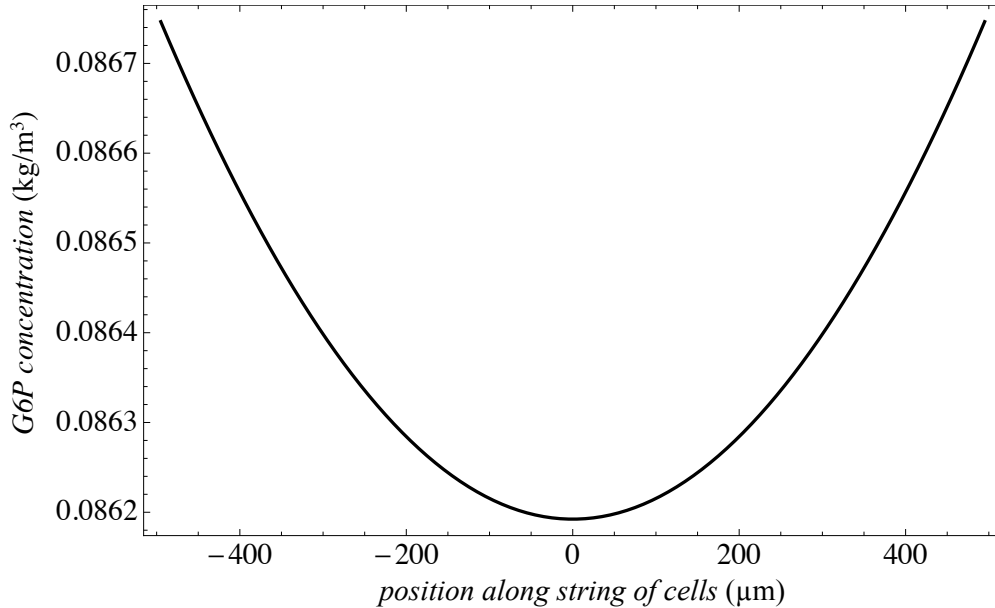
Fig. 9. Plot of G6P concentration along the string of cells at $t = 1000$ s, obtained numerically with the BDF method for the same problem of figure 6, using the differential equation for G6P in [4], solved with the implicit Euler method. The concentration is lowest in the center, and it mirrors a similar curve for glucose concentration.

implicit trapezoidal method, respectively (they are actually the extensions of these algorithms to partial differential equations), and thus – unsurprisingly – they share with their counterparts both advantages and disadvantages. In particular in our tests on this particular diffusion problem we find that:

- both BDF and CN eventually converge to the same equilibrium solution;
- the BDF method requires a larger number of function evaluations to reach the required accuracy;
- the CN method displays unwanted oscillations, similar to those found in the case of the implicit trapezoidal method.

Figure 6 shows contour plots obtained with the BDF and the CN methods, and figure 7 shows the CN solution close to the boundary vs. time: while the global behavior of the solutions is the same, the slow-dying oscillations of the CN method close to the boundary are clearly visible.

The oscillatory behavior of the CN method is due to the reduced damping of the high-frequency spatial modes (see the formula for A, eq. (19)), and it is known to sometime spoil the quality of the CN solutions [23]. These numerical tests indicate that the BDF algorithm – as an extension of the implicit Euler algorithm – is a good choice for an integrator in large scale biophysical simulations.
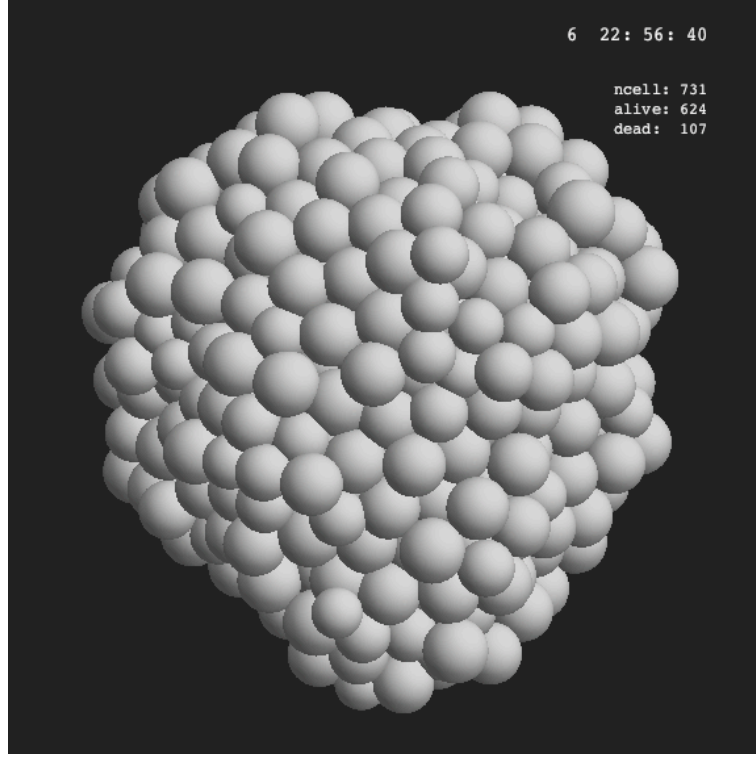
19

Fig. 10. Simulated tumor spheroid in VBL. The simulation started from a single cell and this cluster of cells corresponds to more than 6 days of simulated time.

6 Conclusions

We are developing a simulation program, VBL (Virtual Biophysics Lab) where we aim to include a basic description of biochemical and biomechanical features, to simulate cell clusters, and that should eventually be a numerical model of tumor spheroids, a useful and important *in vitro* model of solid tumors [25]: the stakes are high and this is just one of several attempts to use mathematical and physical methods to understand the biochemical and mechanical properties of tumor spheroids [26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42].

In our modeling effort we proceed in a partly phenomenological way that leads to simple parameterizations: in exchange, we achieve a huge reduction in computational complexity and a considerable reduction of the space-time scale problems that affect simulations aimed at calculating the properties of macroscopic objects starting from microscopic models. We are in an advanced phase of development of the program, and we have already included cell metabolism, growth and proliferation and the extracellular environment. In the previous sections we have discussed the stability properties of the integrators in VBL and in similar programs: we have performed tests on a simple model system and some of the results are interesting on their own, like, e.g., the time evolution and the distribution of G6P in a 1D string of cells (see figures 10 and
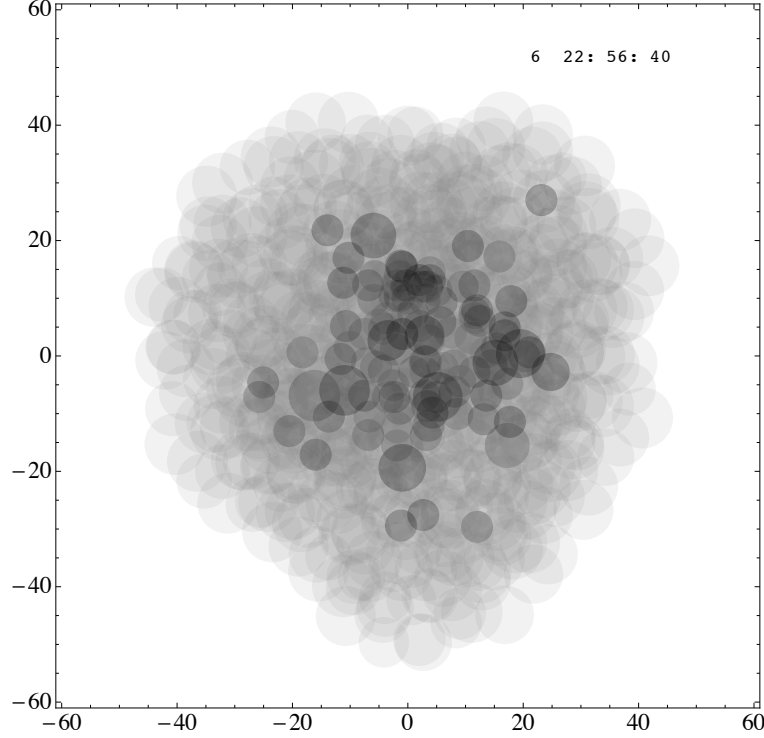
Fig. 11. Distribution of dead cells (dark gray) inside the simulated tumor spheroid of figure 10. Here cells are represented as semitransparent balls, and dead cells in the core become visible. The numbers on the axes are μm.

11). The 3D part of the program, i.e. geometry and biomechanical interactions, is also included in the prototype version of the program, as well as a description of the extracellular microenvironment. We can already simulate large populations of dispersed cells, like those in the culture wells used for in vitro growth, and we have produced numerical estimates that are in excellent qualitative agreement, and in good quantitative agreement, with experimental data [4,5,6]. Figure 10 shows one simulated spheroid and figure 11 is an alternate display that shows the distribution of dead cells inside the same spheroid.

In the present stage of development of the program we are cleaning up the biochemical part of the simulation and building a close integration with the diffusion algorithm to achieve a robust, stable program. Because of discrete – and sometimes random – events in the cells' lives, the evolution is only partly described by differential equations. Moreover – because of cell proliferation – there is a variable number of equations. On the basis of the considerations and the numerical tests described in this paper we have decided to settle on the implicit Euler algorithm and its extension, the Backward Differentiation Formula. The simulations that produced figures 10 and 11 were based on much simpler (and unstable) explicit Euler integration steps, and stability problems showed up at an early stage: soon we shall be able to carry out higher quality

simulations based on the much more stable and reliable methods described here.

Eventually we may have to tackle efficiency and speed as the simulations become larger and more time-consuming. However the choice of BDF is open to improvements in speed, since – as we have seen in the last section – biochemical changes are comparatively slow, and stiffness is brought about mostly by the diffusion terms in the equations: this means that we may eventually use implicit-explicit methods like those described in [43] and this may help us retain the advantages of the implicit BDF algorithm and greatly increase speed – but this is something that we shall study in a future work.

References

[1] A. Cornish-Bowden, Fundamentals of Enzyme Kinetics (Portland Press, London, 2004)

[2] see, e.g., U. Alon, An Introduction to Systems Biology: Design Principles of Biological Circuits (Chapman and Hall/CRC, London, 2007)

[3] R. Chignola and E. Milotti, Physica A 338 (2004) 261

[4] R. Chignola and E. Milotti, Phys. Biol. 2 (2005) 8

[5] R. Chignola, A. Del Fabbro, C. Dalla Pellegrina, and E. Milotti, Phys. Biol. 4 (2007) 114

[6] E. Milotti, R. Chignola, C. Dalla Pellegrina, A. Del Fabbro, M. Farina, D. Liberati, Nuovo Cimento C31 (2008) 109

[7] J. N. Weiss, FASEB J. 11 (1997) 835

[8] R. K. Jain, Cancer Res. 47 (1987) 3039

[9] A. Krol, J. Maresca, M. W. Dewhirst, and F. Yuan, Cancer Res. 59 (1999) 4136

[10] S. Jin, Z. Zador, and A. S. Verkman, Biophys. J. 95 (2008) 1785

[11] K. Groebe, W. Mueller-Klieser, Eur. Biophys. J. 19 (1991) 169

[12] R. A. Medina and G. I. Owen, Biol. Res. 35 (2002) 9

[13] R. Airley et al., Clin. Cancer Res. 7 (2001) 928

[14] H. Heimberg et al, Proc. Natl. Acad. Sci. U.S.A. 93 (1996) 7036

[15] G. Dahlquist, Math. Scand. 4 (1956) 33

[16] J. C. Butcher, Numerical Methods for Ordinary Differential Equations (Wiley, Chichester, 2003)

[17] W. H. Press , S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, Numerical Recipes: The Art of Scientific Computing, 3rd ed. (Cambridge Univ. Press.. New York, 2007)

[18] F. Odeh, IBM J. Res. Develop. 31 (1987) 178

[19] S. Wolfram, The Mathematica Book, fifth ed. (Wolfram Media Inc., Champaign, 2003); see also the online reference http://reference.wolfram.com/mathematica/guide/Mathematica.html.

[20] H. Busch et al, Molecular Systems Biology 4 (2008) 199

[21] I. M. J. J. van de Ven-Lucassen and P. J. A. M. Kerkhof, J. Chem. Eng. Data 44 (1999) 93

[22] N. Gershenfeld, The Nature of Mathematical Modeling, Cambridge Univ. Press, New York, 1999.

[23] D. Britz, O. Østerby, and J. Strutwolf, Comput. Biol. Chem. 27 (2003) 253

[24] Y. Saad, Iterative Methods for Sparse Linear Systems, 2nd ed. (SIAM, Philadelphia, 2003)

[25] R. M. Sutherland, Science 240 (1988) 177

[26] T. Roose, S. J. Chapman and P. K. Maini, SIAM Rev. 49 (2007) 179

[27] R. P. Araujo and D. L. S. McElwain, Bull. Math. Biol. 66 (2004)1039

[28] T. Beyer and M. Meyer-Hermann, Phys. Rev. E 76 (2007) 021929

[29] H. M. Byrne, J. R. King, D. L. S. C. McElwain and L. Preziosi, Appl. Math. Lett. 16 (2003) 567

[30] D. Drasdo, and S. Hoehme, Math. and Comp. Modelling 37 (2003) 1163

[31] D. Drasdo, S. Hoehme, and M. Block, J. Stat. Phys. 128 (2007) 287

[32] J. Galle, G. Aust, G. Schaller, T. Beyer, and D. Drasdo, Cytometry 69A (2006) 704

[33] J. Galle, M. Loeffler, and D. Drasdo, Biophys. J. 88 (2005) 62

[34] Y. Jiang, J. Pjesivac-Grbovic, C. Cantrell, and J. P. Freyer, Biophys. J. 89 (2005) 3884

[35] A. R. Kansal, S. Torquato, G. R. Harsh Iva, E. A. Chiocca, and T. S. Deisboeck, J. Theor. Biol. 203 (2000) 367

[36] F. Meinecke, C. S. Potten, and M. Loeffler, Cell Prolif. 34 (2001) 253

[37] A. A. Patel, E. T. Gawlinski, S. K. Lemieux, and R. A. Gatenby, J. Theor. Biol. 213 (2001) 315

[38] G. Schaller and M. Meyer-Hermann, Phys. Rev. E 71 (2005) 051910

[39] J. A. Sherratt and M. A. J. Chaplain, J. Math. Biol. 43 (2001) 291

[40] G. S. Stamatakos et al., IEEE Trans. Inf. Tech. Biomed. 5 (2001) 279

[41] S. Turner and J. A. Sherratt, J. Theor. Biol. 216 (2002) 85

[42] L. Zhang L., C. A. Athale, and T. S. Deisboeck, J. Theor. Biol. 244 (2007) 96

[43] U. Ascher, S. Ruuth and B. Wetton, SIAM J. Numer. Anal. 32 (1995) 797